Statistical discrimination of uranium ore concentrate using trace element signature: Developing nuclear forensic fingerprint

Liteboho Ntsohi[1], **Iyabo Usman**[1], **Risimati Mavunda**[1,2], **Oscar Kureba**[3] **and Mistura Ajani**[4]

[1]School of Physics, University of the Witwatersrand, Johannesburg 2050, South Africa
[2]Department of Mechanical Engineering, University of Johannesburg, , Johannesburg, South Africa.
[3]Department of Physics and Astronomy, Botswana International University of Science and Technology, Private Bag 16, Palapye, Botswana
[4]Department of Physics with Electronics, University of Ilesa, Ilesa Osun State, Nigeria.

E-mail: liteboho.ntsohi@students.wits.ac.za

Abstract. Effective attribution of nuclear materials intercepted outside regulatory control requires the identification of unique, source inherent characteristics. To date, a limited number of characteristics have been validated as signatures for uranium ore and uranium ore concentrate (UOC), including rare earth element (REE) patterns and trace elements, which serve as geological and geographical indicators. In this study, the concentrations of 11 trace elements, including Ti, V, Ni, Cu, Zn, Ga, Rb, Sr, Y, Zr, Pb, as well as REEs, Th and U in 38 samples from three African regions: Botswana (Southern Africa), Kenya (East Africa), and Nigeria (West Africa) were measured using Inductively Coupled Plasma Mass Spectrometry (ICP-MS). The resulting compositional data were transformed and analyzed using statistical techniques, including Multivariate Analysis of Variance (MANOVA), Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Hierarchical Cluster Analysis (HCA) to test for regional differences, determine the statistical significance of the differences observed, and test the potential discriminative power of individual trace elements. MANOVA revealed statistically significant regional differences in trace element compositions ($p < 0.0001$). LDA identified the elements most responsible for group separation, achieving a high classification accuracy of 91.67% through cross-validation. The first two components in PCA explained 79.15% of the total variance, while HCA further supported the separation of samples into regionally distinct clusters. These results demonstrate the potential of trace element signatures, in combination with multivariate statistical methods, to effectively discriminate uranium samples by region of origin.

1 Introduction

Much like conventional forensics, nuclear forensic science relies on characteristic data and modeling to infer historical information such as origin and production processes [1]. Effective attribution of nuclear materials intercepted outside regulatory control (MORC) is central to nuclear forensic investigations [2]. Uranium Ore Concentrate (UOC) and its precursor, mined ore, are common intermediate front-end products, in the nuclear fuel cycle [3]. UOC is commonly traded between countries and frequently encountered in illicit trafficking incidents [3]. Characteristic parameters in UOC such as chemical composition and most importantly, trace/impurity contents heavily depend on the nature of the raw ore and have geographical and geological signatures [4]. Consequently, forensic attribution of UOC can be based on signatures that persist from ore to UOC. To date, a limited number of characteristic parameters have been validated as forensically significant signatures for mined uranium ore/UOC. Of the signatures examined, the rare earth element (REE) pattern, chondrite normalized patterns in particular, serve as powerful geological and geographical indicators and has become the subject of many studies on attribution of UOC [2, 4]. However, a single signature is sometimes not enough to ascertain the source of MORC; instead, to increase precision, a conglomerate of signatures that complement each other is required to determine the source and, to create a robust nuclear fingerprint [2, 5]. The aim of this study was to demonstrate that measured trace element signature can be used effectively as a tool for geographical source attribution of mined ore/UOC. The concentrations of 27 elements: including 11 trace elements(Ti, V, Ni, Cu, Zn, Ga, Rb, Sr, Y, Zr, Pb), REE (La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu), Th, and U in 38 samples from three African regions: Botswana (Southern Africa), Kenya (East Africa), and Nigeria (West Africa) were measured using Inductively Coupled Plasma Mass Spectroscopy. Comparing the concentrations of multiple trace elements in samples from different countries to determine whether they come from the same country is a complex task [3]. The complexity is brought on by the multidimensional nature of the data, which makes it difficult to visualize the data and recognize patterns. To overcome this multidimensionality, statistical techniques such as DA, PCA and HCA are applied [6, 7]. The data were subjected to a combination of supervised and unsupervised multivariate statistical algorithms such as MANOVA, DA , PCA and HCA; to test whether there are statistically significant differences between samples from different countries, respectively.

2 Materials and Methods

For the purpose of this study, the mined uranium ore samples served as UOC surrogate. Five samples were collected from Botswana, twenty nine from Kenya, and 4 from Nigeria [8, 9]. The samples were crushed, milled and acid digested using deionized water, aquarigia (a 1:3 mixture of HNO_3 and HCl, respectively) and H_2O_2 in preparation for ICPMS analysis [8, 10]. Trace element concentration measurements were then performed using Agilent Technology, 7700 Series ICP-MS. To validate the trace element parameter as a forensically relevant signature, the concentration of 27 elements/variables: Be, B, Sc, Ti, V, Ni, Cu, Zn, Ga, Ge, As, Se, Rb, Sr, Y, Zr, Mo, Cd, In, Sn, Te, Cs, Ba, W, Re, Hg, Tl, Pb, Bi, REEs, Th, and U in the 38 samples were measured. The resulting concentrations in ppb were extracted and prepared for subsequent analysis. The data were subjected to MANOVA, to test whether there are statistically significant differences between samples from different countries, while DA , PCA and HCA were used for classification, dimensionality reduction and data visualization, and clustering, respectively. Data preparation is integral to successful multivariate statistical analysis [11]. Due to the nature of compositional data, it is necessary to transform the data set prior to running multivariate algorithms. Centered log ratio (CLR) transformation was applied to the dataset in CoDAPack software- a software designed to implement suitable methods for compositional data based on Aitchisons' log-ratio methodology [12]. XLSTAT 2025 was used to perform the analysis on the transformed data matrix.

3 Results and discussion

3.1 Multivariate Analysis of variance (MANOVA)

Wilks test and Pillai test were performed in MANOVA (significance level $\alpha = 0.05$) to test if the trace element signature varies significantly between countries. For this, two hypothesis were synthesized; null hypothesis (H_0), and an alternative hypothesis (H_A). H_0 assumes that the variables or the interaction of the corresponding column has no significant effect on the dependent variables (i.e Botswana=Kenya=Nigeria), while H_A assumes that the variable or the interaction of the corresponding column has a significant effect on the dependent variable (atleast one country is different). From the results shown in Table 1, the computed p-value is less than the significance level $\alpha = 0.05$ and $F_{statistic} > F_{critical}$, therefore the H_0 was rejected and H_A was accepted. The risk of rejecting H_0 while it is true was found to be less than 0.01%. It was therefore concluded that atleast one of the countries shows statistically significant differences when comparing the trace element signature. The next section illustrates how the specific elements were identified and determined in driving the variance.

3.2 Unidimensional test for equality of means

Unidimensional test for equality of means is a preliminary statistical test used in DA that is based on Analysis of variance(ANOVA). The test uses a hypothesis similar to that described in MANOVA for each variable separately, helping to identify individual elements that contribute significantly to group differences [13]. In addition to the p-value and the f-value, the test also evaluates the value of lambda for each element. Lambda is a measure of the proportion of variance in dependent variables unaccounted for by differences in levels of the independent variable (grouping variable) [14]. A lambda value of zero indicates that there exists no variance not explained by the independent variable (which is ideal). The closer the lambda is to zero, the more the variable/element in question contributes to the model. In combination with p-value < α and $F_{value} > F_{sig}$ a null hypothesis can be rejected.

	Samples	
	Wilks' test	Pillai's Test
F Observed values	63.609	43.178
DF1	54	54
DF2	48	50
F Critical value	1.599	1.589
p-value	<0.0001	<0.0001

Table 1: One-way MANOVA statistics for uranium ore samples using trace element signature

Figure 1 is a graphical representation of lambda result of the test. Elements with lambda > 0.5 do not contribute significantly to variance and were not selected for subsequent PCA and HCA.

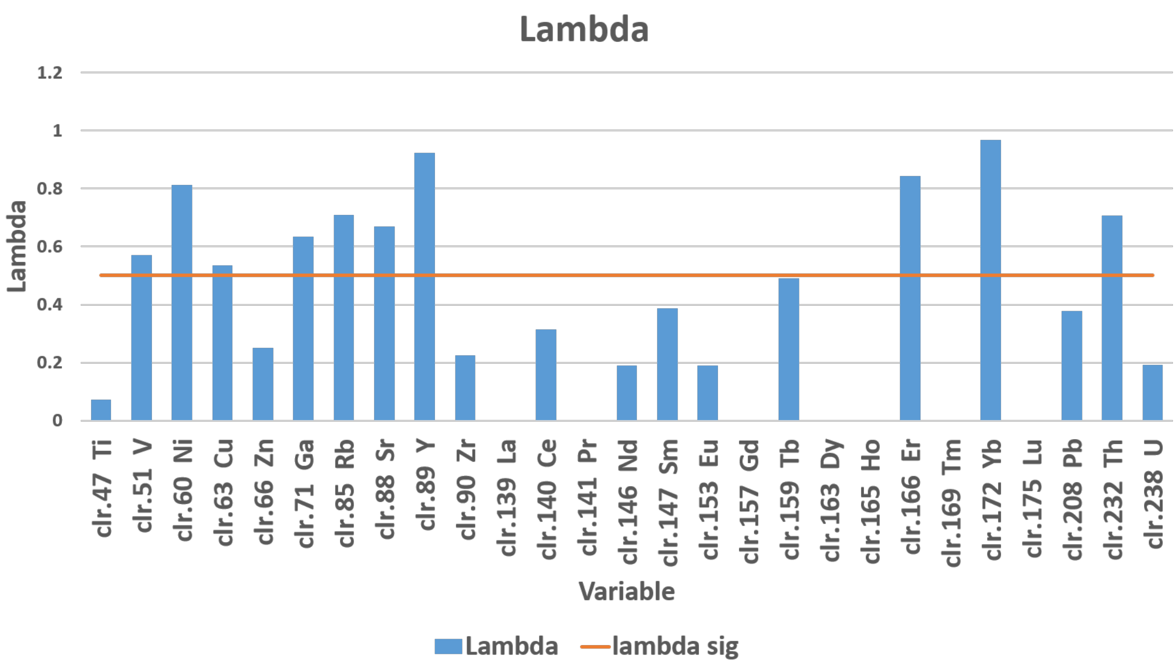

Figure 1: Bar graph showing calculated lambda for the individual elements

3.3 Discriminant Analysis (DA)

DA was used for its pattern recognition capability, which was used to build a classification model from trace element data. DA generated discriminant functions by maximizing the between-class variance to within class variance ratio to ensure high efficiency of the classification model [1]. The prediction/classification ability of the model was evaluated using cross-validation [1]. One random sample was removed per country from the dataset to form the test set, while the remaining n-3 samples were used to train the model. Once training was complete, the model was used to predict the origin of the test set. Due to the relatively small data size and the unequal distribution of the samples from each country, the process was repeated so that 30% of the largest group(Kenya) is used to test the accuracy of the model. The procedure was therefore repeated nine times. The performance of the model can be seen in Table 2. The results show that samples from Botswana and Kenya were correctly classified all the time; however, on two occasions a sample from Nigeria was misclassified as Botswana. The developed model had a prediction ability of 91%, which is acceptable.

	from/to	Botswana	Kenya	Nigeria	% correct
Training Sample	Botswana	9	0	0	100%
	Kenya	0	9	0	100%
	Nigeria	0	0	9	100%
	Recognition ability				100%
Cross-validation	Botswana	9	0	0	100%
	Kenya	0	9	0	100%
	Nigeria	2	0	7	75%
	Prediction ability				91.67%

Table 2: Confusion matrix for training and cross-validation results based on trace element concentrations

3.4 Principal Component Analysis (PCA)

To reveal whether the samples form natural clusters based on their trace element parameters, unsupervised methods; PCA and HCA were applied. PCA is a dimensionality reduction tool which transforms data from a multidimensional space and enables visualization in fewer dimensions (PCs) while preserving the most important variations. PCs are linear combinations of the original data. To determine the number of significant PCs, eigenvalues were examined. Applying the Kaiser-Guttmann rule, only principal components/factors with eigenvalue greater than 1 were considered significant dimensions. The first two PCs, F1 and F2, met the criteria and accounted for 79.15% of the total variance. These were used to construct the PC biplot shown in Figure 2. The left-hand side of Figure 2 shows the results of the Kaisers varimax orthogonal rotation. The rotation moves the component axis so that the projections of each variable (trace element) onto the factor axes are either near 1 or the origin. The position of the variable is indicative of significance. However, due to the removal of insignificant variables prior to running PCA, all elements are closer to 1 than to the origin. This further corroborates the findings of unidimensional tests in DA. The right-hand side of Figure 2 shows the clustering that is formed. Three samples from Nigeria create a group high on the F2 axis; however, one sample, Nig1, is detached from the group. Samples from Botswana also form a distinct group, and the same is observed for samples from Kenya with the exception of three samples (MS14, MS15, MS8) that are detached from the group. These results are in agreement with those obtained during DA.

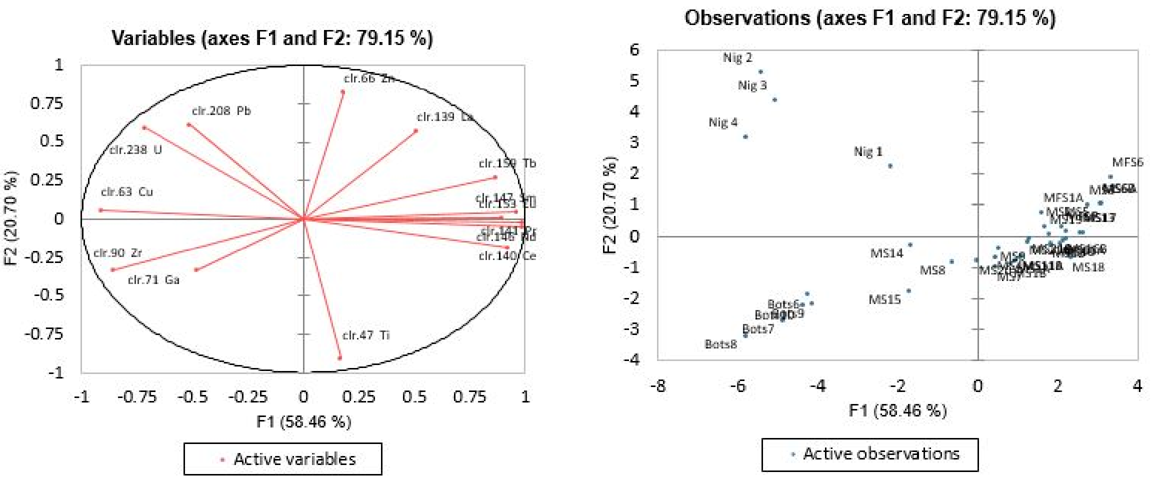

Figure 2: (LHS: PC loading biplot of trace elements in ore samples. RHS: PC score plot illustrating the differentiation between ore samples according to their origin

3.5 Hierarchical Cluster Analysis (HCA)

Unlike DA, HCA does not assign a sample to a class (country), rather it separates the samples into clusters so that the level of association is strong between samples in the same cluster and weak between samples in different clusters [1]. This was achieved using Wards' method with Euclidean distance to calculate the sample interpoint distance. HCA was performed on uranium ore samples with trace element parameters as input variables. The generated dendogram is shown in Figure 3 with a dissimilarity threshold of 52. The results indicate 5 distinct clusters and one subgroup. C1 contained 5 samples labeled Bots6, Bots7, Bots8, Bots9 and Bots10 which are all from Botswana. C2 is a cluster of 1 sample labeled Nig 1 while C3 contains samples labeled Nig 2, Nig 3, Nig 4. Both clusters contain samples from Nigeria. The same can be observed with C4 and C5 which contain samples from Kenya only. The separation of samples from similar countries into distinct clusters is evidence of the discriminative ability of the trace element signature. This shows that the signature has the ability not only to separate ores on the basis of country of origin but also to distinguish between samples for different regions within the country, while still recognizing that the samples are related. The results obtained support the findings of PCA and DA, showing regionally consistent clusters.

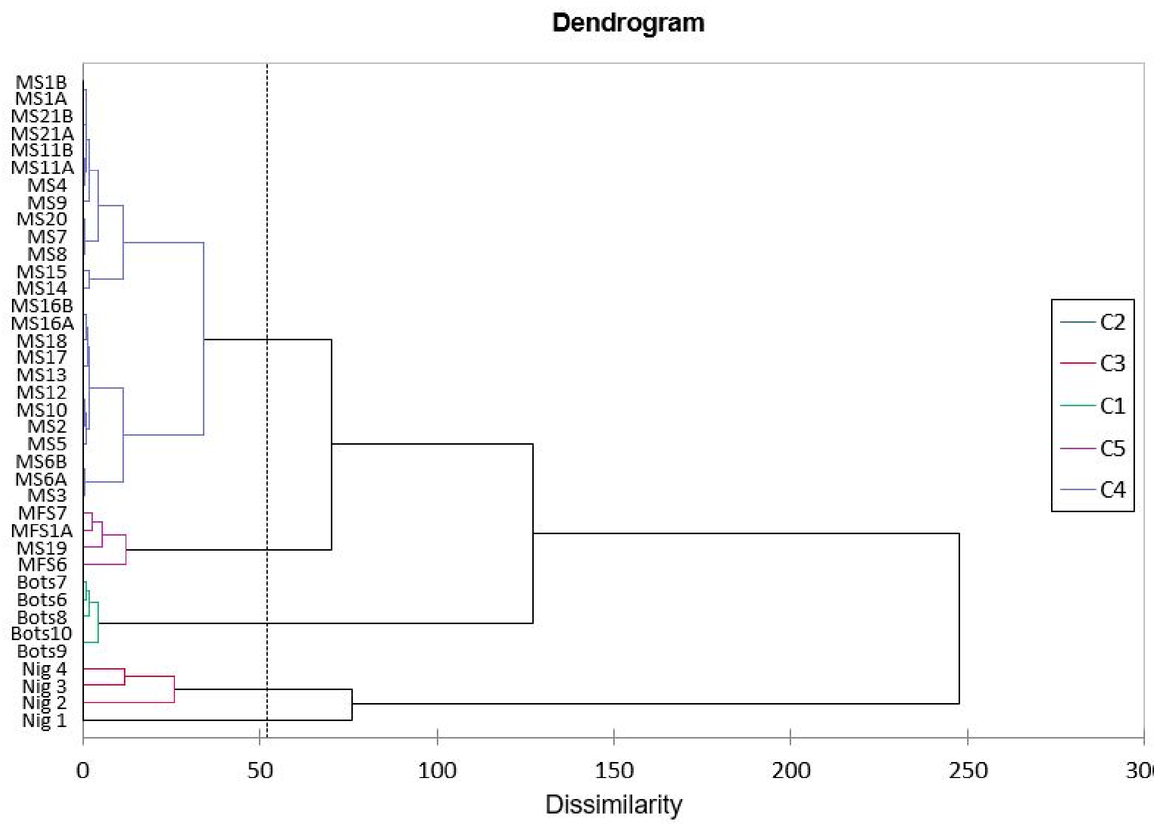

Figure 3: HCA dendrogram of uranium ore samples from three regions based on trace element concentrations

4 Conclusion

This study demonstrated that the trace element signature is forensically significant and could be applied in combination with REE signature to discriminate uranium ore/UOC samples from various countries and regions throughout Africa for source attribution purposes. By using multiple multivariate statistical techniques (MANOVA, DA, PCA, and HCA), the study showed how these techniques were effective tools for signature validation and interpretation for source attribution. MANOVA revealed statistically significant differences among countries/regions ($F_{statistic} > F_{critical}$, $p < 0.05$), indicating that trace element signature varies by source. Unidimensional test used in DA identified Ti, Zn, Zr, Ce, Pr, Nd, Eu and Pd as key elements driving the differences between samples of different origins. The classification model developed using trace element parameters was found to have predictive accuracy of 91.76%. Further demonstrating the strength of the trace element signature for classification. Results from PCA and HCA also demonstrated that trace element signature could be used to distinguish samples of unknown origin. The application of both tools resulted in clusters that were consistent with the sampling locations. Although the results revealed that the trace element signature is a good geographical signature, the sample size was small and the sample sizes were not balanced between the countries. To strengthen these findings, extensive research is required on a larger set of uranium ore samples from sources across the continent. In addition, incorporating machine learning techniques would improve classification and prediction accuracy and assist in developing a comprehensive continental reference database for nuclear forensic applications.

5 Acknowledgement

This work was supported by AFRETEC Network by Carnegie Mellon Africa Funds.

References

[1] X. Shao, W. Bu, Y. Ni, H. Wang, X. Liu, C. Yang, and F. Hao, "Origin assessment of uranium ores using multivariate statistical method based on their rare-earth elemental parameters," *Nuclear Analysis*, vol. 1, no. 3, p. 100027, 2022.

[2] L. Ghimire and E. Waller, "The role of nuclear forensics for determining the origin of nuclear materials out of regulatory control and nuclear security," *International Journal of Nuclear Security*, vol. 7, no. 1, p. 10, 2022.

[3] M. Lin, Y. Zhao, L. Zhao, L. Li, F. Wang, L. Zhu, X. Hu, and W. Ning, "Tracing origins of uranium ore concentrates (uocs) by multidimensional statistical analysis of rare-earth impurities," *Journal of Analytical Atomic Spectrometry*, vol. 30, no. 2, pp. 396–402, 2015.

[4] V. Badaut, M. Wallenius, and K. Mayer, "Anion analysis in uranium ore concentrates by ion chromatography," *Journal of radioanalytical and nuclear chemistry*, vol. 280, no. 1, pp. 57–61, 2009.

[5] E. Keegan, M. Wallenius, K. Mayer, Z. Varga, and G. Rasmussen, "Attribution of uranium ore concentrates using elemental and anionic data," *Applied Geochemistry*, vol. 27, no. 8, pp. 1600–1609, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0883292712001321

[6] Q. T. Nguyen, T. T. Nguyen, V. N. Le, N. T. Nguyen, N. M. Truong, M. T. Hoang, T. P. T. Pham, and Q. M. Bui, "Towards a standardized approach for the geographical traceability of plant foods using inductively coupled plasma mass spectrometry (icp-ms) and principal component analysis (pca)," *Foods*, vol. 12, no. 9, 2023. [Online]. Available: https://www.mdpi.com/2304-8158/12/9/1848

[7] D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira, and R. M. Maggio, "Use of principal component analysis (pca) and hierarchical cluster analysis (hca) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective," *Trends in Food Science Technology*, vol. 72, pp. 83–90, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924224417306362

[8] L. Ntsohi, I. Usman, R. Mavunda, and O. Kureba, "Characterization of uranium in soil samples from a prospective uranium mining in serule, botswana for nuclear forensic application," *Journal of Radiation Research and Applied Sciences*, vol. 14, no. 1, pp. 23–33, 2021.

[9] P. Kilavi, M. Kaniu, J. Patel, and I. Usman, "Assessment of heavy metal pollution in soil and associated risks in the environs adjacent to a heavy mineral sand mine on the south coast of kenya," *Water, Air, & Soil Pollution*, vol. 234, no. 12, p. 748, 2023.

[10] S. O. John and I. T. Usman, "Isotopic profiling of natural uranium mined from northern nigeria for nuclear forensic application." *South African Journal of Science*, vol. 118, no. 3-4, 2022.

[11] S.-L. J, "Development and validation of a method for origin determination of uranium-bearing material - jrc-itu-tn-2008/25," Ph.D. dissertation, Vilnius University, Vilnius (Lithuania), Karlsruhe (Germany), 2008.

[12] S. Thió-Henestrosa and J. A. Martín-Fernández, "Detailed guide to codapack: a freeware compositional software," in *Compositional Data Analysis in the Geosciences: From Theory to Practice*. Geological Society of London, 01 2006. [Online]. Available: https://doi.org/10.1144/GSL.SP.2006.264.01.08

[13] B. G. Tabachnick and L. S. Fidell, *Multivariate Analysis of Variance (MANOVA)*. Springer Berlin Heidelberg, 2025, p. 1616–1619. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-69359-9_398

[14] P. Lovie, "Book reviews : Everitt BS, dunn G 1991: Applied multivariate data analysis," *Statistical Methods in Medical Research*, vol. 0, pp. 114–115, 2016.